

# Apport de la sélection de primitives point à un suivi robuste

## Improving Robust Feature Tracking by Selection

Chafik Kermad<sup>1</sup>

Christophe Collewet<sup>1</sup>

Kacem Chehdi<sup>2</sup>

<sup>1</sup> Cemagref, Rennes

<sup>2</sup> Enssat-Lasti, Lannion

Cemagref, 17, Avenue du Cucillé, CS 64427, 35044 Rennes  
chafik.kermad@cemagref.fr

### Résumé

*Ce papier traite de la sélection automatique de points d'intérêt en vue de leur suivi dans une séquence d'images et ce dans un contexte de vision dynamique ou d'asservissement visuel. La sélection permet, dès l'image initiale, de proposer des candidats susceptibles d'être correctement appariés le long de cette séquence grâce à un algorithme de suivi. Pour ce faire, à partir d'un détecteur classique de points d'intérêt, une sélection de points situés dans des zones optimisant un critère d'entropie est effectuée. Des résultats sur diverses séquences d'images mettent en évidence l'intérêt de l'approche proposée quant au suivi effectif des points sélectionnés.*

### Mots Clef

Extraction de points d'intérêt, Sélection de points d'intérêt, Entropie, Images complexes, Suivi de points.

### Abstract

*This paper deals with the automatic selection of points of interest for their tracking along a sequence of images in a context of dynamic vision or visual servoing. The selection allows, from the first frame, to propose candidates having a high confidence index to be correctly matched along the sequence. Starting from a classical detector of points of interest, a selection of points located in areas optimizing an entropy criterion is carried out. Various experiments show the interest of the proposed approach for the effective tracking of the selected points.*

### Keywords

Interest points extraction, Interest points selection, Entropy, Complex images, Feature points tracking.

## 1 Introduction

Le problème de l'appariement ou de la mise en correspondance de primitives entre deux images reste un problème ouvert en vision par ordinateur. C'est en effet un problème

difficile dont peuvent dépendre, suivant les approches utilisées, certaines tâches de plus haut niveau comme par exemple le calcul du mouvement 2D, le suivi de cible, la vision dynamique (active ou non), ou l'asservissement visuel. Ce problème devient plus complexe encore dès lors que des amers visuels ne peuvent être utilisés. Citons par exemple le cas d'applications dans les domaines de la chirurgie, de l'agro-alimentaire, de l'agriculture ou celles intervenant en environnement inconnu comme les applications sous-marines ou spatiales.

Dans ce contexte, parmi les primitives facilement et rapidement extractibles dans une séquence d'images, seuls des *points d'intérêt* sont susceptibles d'être présents dans une image complexe comme c'est le cas par exemple dans [5, 7, 8, 21]. Classiquement la façon de faire se décompose alors en trois phases : une première phase extrait les points d'intérêt, vient ensuite une phase d'appariement de ces points d'une image à l'autre, puis, finalement, une phase de validation ou de réjection éventuelle de cet appariement (voir [31] pour un exemple complet). Dans le cas où les primitives point constituent les informations visuelles nécessaires à l'élaboration d'une loi de commande, comme c'est le cas en asservissement visuel ou en vision active, cette façon de faire n'est pas du tout adaptée puisqu'un mauvais appariement de ce type de point conduit invariablement à la mise en échec de la tâche à réaliser. Pour pallier à ce problème nous proposons, après une extraction classique de points d'intérêt, de sélectionner les points à suivre dès la première image au sens où ils seront correctement suivis tout au long de la séquence. Très peu d'auteurs se sont intéressés à ce problème. En effet, dans un passé récent ce problème était moins présent puisqu'alors les scènes observées étaient principalement des scènes d'intérieur où il s'agissait d'extraire des coins ou des jonctions, primitives pour lesquelles sont adaptés les détecteurs de points d'intérêt. Toutefois, Tomasi et Kanade [30], puis Shi et Tomasi [28], ont proposé un classement des points extraits suivant le conditionnement, bon ou mauvais, d'une ma-

trice utilisée lors du suivi. Elle a été obtenue en considérant comme modèle de déplacement une translation pure entre deux images successives [22] (cette approche sera détaillée en 3). Très récemment, sur des images complexes sous-marines (images de fumerolles), F-X. Espiau a proposé une approche pyramidale permettant de ne conserver à la résolution la plus haute que les points les plus robustes suivant un critère d'appariement de points extraits à plusieurs niveaux de la pyramide [8].

Dans notre cas, nous nous intéressons plus particulièrement à l'analyse de séquences d'objets naturels rigides, de formes complexes, non structurés et ayant un caractère texturé dont on n'a aucune connaissance *a priori*. Notre approche est basée sur la mesure de l'entropie de cooccurrence des niveaux de gris de l'image initiale. Elle traduit le fait qu'un point d'intérêt se situe dans une zone riche en information ou non, et, par conséquent, qu'il sera plus à même d'être apparié pour un algorithme classique de suivi. Ces algorithmes sont en effet le plus souvent basés sur une mesure de similarité d'une fenêtre d'analyse entre deux images successives comme on le verra en section 2.

Cet article est structuré comme suit. Tout d'abord, nous présentons dans la section 2 un rapide panorama des méthodes d'extraction et de suivi de points d'intérêt rencontrés dans la littérature. Dans un but de comparaison avec notre approche, nous détaillons ensuite, en section 3, la méthode d'extraction et de sélection des points d'intérêt décrite dans [30]. Nous rappelons ensuite, en section 4, la méthode proposée par Soatto *et al.* dans [29] qui nous permettra de réaliser le suivi des points sélectionnés par notre approche, approche exposée en section 5. Des résultats expérimentaux sont ensuite présentés en section 6. Enfin, en section 7, une conclusion rappelle les principaux résultats obtenus et évoque quelques perspectives de recherche.

## 2 État de l'art

Un des problèmes majeurs dans la réalisation d'un système de vision/robotique est l'extraction et le suivi d'indices visuels significatifs dans les images. L'objet de cette section est de faire un bref état de l'art sur les différentes approches proposées pour traiter ces deux problèmes.

### 2.1 Extraction de points d'intérêt

Comme rappelé en introduction, les points d'intérêt s'avèrent posséder le caractère le plus général. Ils correspondent généralement à un changement bidimensionnel du signal d'intensité comme par exemple les coins, les jonctions mais aussi d'autres points de l'image. De nombreuses méthodes pour l'extraction de tels points sont proposées dans la littérature [26, 27, 32]. Malgré leur diversité, elles peuvent être classées en deux grandes catégories:

- La première extrait les points d'intérêt à partir de primitives plus évoluées comme les contours «*Contour-based methods*». Citons par exemple [1] où les changements de courbure sont exploités pour l'extraction de coins ou de terminaisons, [18] et [3] pour ce qui

concerne les extrema de courbure et [23] pour ce qui est des passages à zéro. Cependant, pour pouvoir considérer ce type de point comme une information visuelle valide, il nous faut nous assurer que la propriété choisie (changement, extrema ou zéro de courbure) se conserve par projection perspective. Par exemple, les extrema de courbure utilisés pour la mise en correspondance en stéréovision, comme dans [18] ou [3], ne sont conservés que pour une petite région de l'espace<sup>1</sup>. Par ailleurs, on peut montrer [4], que dans les cas non dégénérés (c'est-à-dire quand l'image de l'objet plan n'est pas une droite), les points de courbure nulle ou points d'inflexion sur l'objet sont aussi des points d'inflexion dans l'image, certains auteurs utilisent cette propriété [23]. De la même façon, on peut montrer que les points stationnaires sont conservés par projection perspective [4].

- La seconde catégorie de méthodes se propose d'extraire les points directement à partir du signal d'intensité selon une approche différentielle «*Signal-based methods*» ou en approximant les points recherchés avec des modèles théoriques «*Template fitting methods*». Les méthodes basées sur l'intensité lumineuse font l'objet de nombreux travaux [12, 19, 22, 24, 26, 27, 30]. Toutefois, pour que ces points puissent être considérés comme des points caractéristiques, il nous faut vérifier l'invariance de cette intensité par transformation perspective. En fait, seuls des objets caractérisés par une *surface lambertienne* vérifient cette propriété. Néanmoins, il existe d'autres types de surface pour lesquelles l'intensité lumineuse varie peu en fonction de la position de la caméra [13, 14]. Cette contrainte est donc très liée à la nature de l'objet.

### 2.2 Suivi de points d'intérêt

La technique de suivi de points la plus répandue consiste à minimiser une mesure de corrélation des intensités de deux images successives (technique dénommée SSD pour «*Sum of Squared Differences*»). Dans [30], Tomasi et Kanade ont présenté un algorithme de suivi basé sur cette mesure. Les auteurs exploitent la continuité temporelle du mouvement et supposent que la déformation géométrique d'une image à l'autre correspond à une translation pure. Sous ces hypothèses et lorsque la fréquence d'acquisition des images est élevée, cette méthode fonctionne correctement. Elle se révèle néanmoins moins robuste dans le cas où les images ont subi des rotations ou des changements d'échelle importants ou lorsqu'une forte disparité d'une image à l'autre existe. Par la suite, afin de prendre en compte des mouvements plus complexes, Shi et Tomasi [28] ont proposé un modèle de déformation affine et non plus une simple translation. Pour examiner la robustesse des primitives extraites et évaluer la quantité d'information

1. Généralement, en stéréovision, les deux caméras sont suffisamment proches les unes des autres pour que les extrema de courbure correspondent bien, dans les deux images, au même point sur l'objet.

présente autour d'un point, les auteurs proposent d'analyser les valeurs propres d'une matrice d'auto-corrélation calculée localement. Seuls les points les plus discriminants suivant ces valeurs sont candidats au suivi. Une amélioration de cette approche a par la suite été proposée dans [31] pour traiter le cas des mauvais appariement. Pour ce faire, les auteurs utilisent une règle de rejet basée sur une comparaison des résidus calculés entre la première image de la séquence et l'image en cours de traitement. Cependant, on peut noter que cette approche [28, 30] présente une forte sensibilité aux conditions d'éclairage et aux bruits. De petites modifications du contenu de l'image dues à des transformations photométriques: contraste (changement d'illumination de la scène), occultation (ajout et suppression de pixels), bruit du capteur, influent sur la précision du résultat d'appariement entre les indices visuels.

Pour remédier au problème de sensibilité des conditions d'éclairage, Hager et Belhumeur décrivent dans [10] un algorithme basé également sur la minimisation d'une SSD entre un modèle d'apparence, décomposé sur une base de vecteurs propres représentant l'illumination et l'image courante. Ces travaux ont conduit à des méthodes performantes et robustes mais impliquent une connaissance *a priori* sur la nature et l'apparence des objets observés. Plus récemment, sans recours à ce type de connaissance, Soatto *et al.* ont présenté dans [29] une extension de l'algorithme de Shi-Tomasi-Kanade en prenant à la fois compte des sources de variation de la géométrie et de l'illumination dans la séquence analysée. C'est cette dernière méthode qui nous permettra de suivre les points sélectionnés par notre approche. Cette méthode sera brièvement exposée en 4.

Sans être exhaustif, nous pouvons également citer les approches de type «Multiple Hypothesis Tracking» et «Conditional Propagation» proposées respectivement dans [6] et [15]. Bien qu'efficaces, ces approches impliquent un coût calculatoire non négligeable qui, dans un contexte d'asservissement visuel ou de vision active, peut être pénalisant.

### 3 La méthode de Tomasi et Kanade

Dans un but de comparaison avec notre approche, nous rappelons, ici, la méthode proposée dans [30], tant pour l'extraction des points d'intérêt que pour leur sélection. Cette méthode repose sur l'hypothèse générale où l'intensité lumineuse ne varie pas pendant l'acquisition des images. Dans ces conditions, en notant  $\mathbf{x}$  un point d'une fenêtre d'analyse  $W$  de l'image  $I(\mathbf{x}, t)$  acquise à l'instant  $t$ , la déformation de cette image peut être décrite par la relation suivante :

$$I(\mathbf{x}, t) = I(\delta(\mathbf{x}), t + \tau) \quad \forall \mathbf{x} \in W \quad (1)$$

où  $\delta(\cdot)$  exprime la transformation linéaire des coordonnées du point  $\mathbf{x}$  permettant le passage de l'image acquise à l'instant  $t$  à celle acquise à l'instant  $t + \tau$ . Dans un premier temps cette transformation est supposée être une translation pure  $d$ , d'où :

$$\delta(\mathbf{x}) = \mathbf{x} + d \quad (2)$$

Cependant, la présence de bruit et les approximations du modèle de mouvement font que l'équation (2) n'est généralement pas satisfaite. Il est alors plus judicieux de considérer ce problème dans le cadre d'une optimisation où il faut trouver le paramètre  $d$  qui minimise le résidu suivant :

$$\epsilon = \sum_W (I(\mathbf{x} + d, t + \tau) - I(\mathbf{x}, t))^2 \quad (3)$$

Pour effectuer cette minimisation, on considère des faibles déplacements et une acquisition suffisamment rapide des images. Le modèle de l'intensité est alors approximé en utilisant un développement en série du Taylor de premier ordre. Il vient alors :

$$\epsilon \approx \sum_W (\nabla I(\mathbf{x}, t)^T d + I_t(\mathbf{x}, t)\tau)^2 \quad (4)$$

où  $\nabla I$  et  $I_t$  sont respectivement les gradients spatial et temporel de l'image calculés au point  $\mathbf{x}$ . Par suite, après différenciation de (4) par rapport à  $d$  et réécriture sous forme matricielle, on obtient le système linéaire suivant :

$$Gd = z \quad (5)$$

$$\text{où } G = \sum_W \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix} \quad \text{et } z = -\tau \sum_W I_t \nabla I$$

Finalement,  $d$  est obtenu grâce à (5) en utilisant un schéma itératif de type Newton-Raphson pour tenir compte de l'approximation introduite par le développement en série de Taylor.

Dans ces conditions un point d'intérêt est considéré comme bon candidat au suivi si, simplement, la résolution de (5) est stable, soit en d'autres termes, si la matrice  $G$  est bien conditionnée. De plus, les auteurs caractérisent le point considéré suivant les valeurs propres  $\lambda_1$  et  $\lambda_2$  de  $G$  et, finalement, l'accepte si :

$$\min(\lambda_1, \lambda_2) > \lambda \quad (6)$$

où  $\lambda$  est un seuil à définir par l'utilisateur en fonction de la scène observée.

### 4 Algorithme de suivi utilisé

Nous rappelons maintenant l'algorithme de suivi que nous utiliserons pour valider notre approche. Il s'agit de l'algorithme proposée par Soatto *et al.* dans [29].

Le modèle de transformation défini par la relation (1) présente l'avantage d'être à la fois simple et assez général. Néanmoins, dans la pratique, les conditions d'acquisition des images ne peuvent pas toujours être contrôlées et, de ce fait, l'hypothèse d'invariance de l'intensité lumineuse n'est plus valide. Dans ce contexte, les auteurs proposent de modifier les modèles décrits dans [28, 30] (translation  $d$  et transformation affine  $A$ ) en prenant en compte les changements de contraste et d'intensité lumineuse par l'adjonction respectivement des termes  $\mu$  et  $\xi$ . Par suite, comme en 3, il s'agit de trouver  $A$ ,  $d$ ,  $\mu$  et  $\xi$  minimisant :

$$\epsilon = \sum_W (\mu I(A\mathbf{x} + d, t + \tau) + \xi - I(\mathbf{x}, t))^2 w(\mathbf{x}) \quad (7)$$

$w(\cdot)$  étant une fonction de pondération gaussienne.

L'obtention effective des paramètres recherchés suit alors la même démarche que celle proposée dans [28, 30], une décomposition en série de Taylor puis un schéma itératif de type Newton-Raphson permettent d'aboutir.

## 5 Approche de sélection proposée

Dans cette section, nous décrivons la méthode proposée pour la sélection des points d'intérêt, extraits ici suivant la méthode décrite en 3, en vue de leur suivi par l'algorithme rappelé en 4.

Intuitivement on conçoit assez bien qu'un point se trouvant dans une zone de niveau de gris homogène a beaucoup moins de chance d'être bien apparié et suivi qu'un point se trouvant dans une zone riche en information. *A contrario*, plus la texture présente autour d'un point est riche et variée, plus l'information image est discriminante. Pour évaluer la quantité d'information présente autour d'un point, la théorie de l'information offre des outils probabilistes basés sur des mesures entropiques. En analyse d'images, ces mesures sont, entre autres, utilisées en codage, en compression ou en caractérisation et segmentation de textures [2, 9, 25, 33]. Néanmoins, à notre connaissance, elles n'ont pas été appliquées à l'extraction et à la sélection de points d'intérêt. D'une façon générale, la mesure d'entropie permet de donner des indications sur l'ordre, la richesse, le comportement statistique et la complexité des configurations locales des pixels d'une image [11]. Dans cette optique, nous proposons une méthode basée sur la mesure des entropies de cooccurrences des niveaux de gris de l'image pour sélectionner, parmi les points extraits suivant le critère (6) décrit en 3, ceux qui sont considérés par notre approche comme de bons candidats pour le suivi.

Pour ce faire, considérons une image  $I$ , codée sur un ensemble de  $N_{ng}$  niveaux de gris et étant donné un vecteur de déplacement  $(\Delta_x, \Delta_y)^T$ . Pour deux niveaux de gris  $k$  et  $l$ , la valeur à la position  $(k, l)$  de la matrice  $C$  de cooccurrence est donnée par le nombre de couples  $(x, y)$  de pixels tels que  $I(x, y) = k$  et  $I(x + \Delta_x, y + \Delta_y) = l$ . Pratiquement,  $C$  est calculée pour les déplacements  $(\Delta_x, \Delta_y) \in \{(1, 0), (1, 1), (0, 1), (-1, 1)\}$ . Afin de ne pas utiliser les  $N_{ng}$  niveaux gris de l'image et réduire ainsi le coût de calcul, nous simplifions cette image en limitant les niveaux gris en un nombre  $G$ , suffisant pour évaluer les probabilités de cooccurrences. Cette opération est réalisée en utilisant une technique automatique de multi-seuillage [17]. La valeur de l'entropie sur une fenêtre d'analyse de taille  $W_E^2$  est alors donnée par :

$$\text{Ent} = - \sum_{k \in G} \sum_{l \in G} p_{kl} \log(p_{kl}) \quad (8)$$

où

$$p_{kl} = \frac{C[k, l]}{W_E^2} \quad (9)$$

On peut alors montrer que l'exploitation du maximum d'entropie conduit à une distribution uniforme des probabilités

d'occurrence<sup>2</sup> [16, 20].

L'application de ce critère sur une image se concrétise par la localisation de fenêtres texturées présentant des configurations composées d'un nombre de niveaux de gris distribués de manière équitable. Ceci se traduit principalement par la détection de zones comportant des jonctions multiples, par exemple de type 'Y', 'T', '+', avec invariance par rapport aux rotations (modulo le calcul de  $C$  suivant  $(\Delta_x, \Delta_y)^T$ ), aux changements d'intensité et aux facteurs d'échelles. Ce critère est ensuite appliqué sur le résultat du détecteur de points d'intérêt, d'une part pour les localiser précisément dans la fenêtre et, d'autre part, pour les classer suivant la valeur d'entropie. Seuls les points se trouvant dans des fenêtres ayant une valeur d'entropie supérieure à un seuil  $s_e$ , fonction de toutes les entropies calculées, sont retenus pour le suivi.

Notons enfin que l'utilisation des cooccurrences plutôt que des occurrences simples permet d'obtenir des estimations d'entropies robustes, au sens de la stabilité, toute en tenant compte des relations spatiales entre les pixels.

## 6 Résultats expérimentaux

Pour comparer et évaluer quantitativement l'approche développée, nous avons appliqué l'algorithme de suivi, décrit en 4, en utilisant une première version où la sélection et le classement des points d'intérêt sont faits suivant le seul critère lié aux valeurs propres (Méthode VP, décrite en 3) et une seconde version où la sélection et le classement des points sont faits suivant le critère d'entropie (Méthode ENT). Pour la méthode VP, un tri des points d'intérêt suivant les valeurs propres locales est effectué et les  $N_{b_{Sel}}$  points les mieux classés sont sélectionnés pour le suivi ;  $N_{b_{Sel}}$  étant le nombre de points issus de la sélection par la méthode ENT.

L'ensemble des résultats présentés dans la suite ont été obtenus dans les mêmes conditions expérimentales : le seuil appliqué aux valeurs propres (paramètre  $\lambda$ ) est égal à 0.1, la taille de la fenêtre  $W$  utilisée dans le suivi est égale à 9 et le seuil sur l'entropie  $s_e$  est égal à  $\frac{3}{2}\mu_e$ , où  $\mu_e$  représente la valeur moyenne des entropies locales calculées sur des fenêtres de taille  $W_E = 16$ . Afin de conserver une localisation précise des points lors du suivi, nous avons introduit un seuil sur les résidus donnés par (7) : seuls les points ayant un résidu inférieur à un seuil normalisé  $s_e$  sont suivis. En pratique, une valeur de 0.05 convient.

Nous présentons maintenant, quatre expérimentations complémentaires. Une première appliquée à une séquence d'images d'objets complexes à caractère plus ou moins texturé, une seconde sur une scène d'images synthétiques, une troisième sur des images composées de régions homogènes, et, finalement, une dernière sur une scène d'images composées de régions fortement texturées.

2. En effet, un calcul simple permet de vérifier que (8) est maximum pour  $\hat{p}_{kl} = W_E^2 / N_g$ ;  $N_g$  étant le nombre des niveaux de gris présents dans la fenêtre d'analyse.

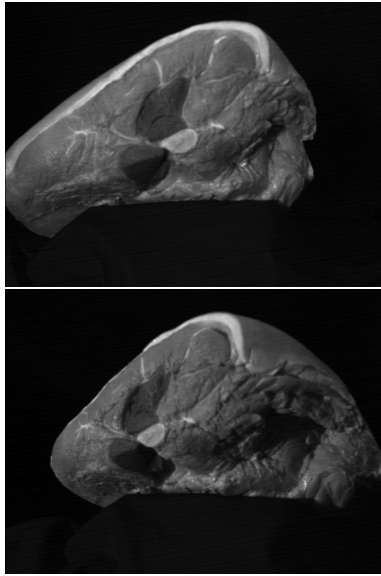


FIG. 1 – Images issues de la séquence PA.

#### Séquence sur un objet complexe

L'algorithme développé a été testé sur plusieurs séquences réelles d'objets complexes. Nous présentons ici quelques résultats obtenus sur une séquence issue de l'observation d'un produit alimentaire (séquence PA, figure 1).

Les 25 images de la séquence ont été acquises par une caméra CCD fixée sur un robot (plate-forme décrite dans [4]). Dans cette expérience, la caméra tourne de gauche à droite par rapport à l'objet et pendant la rotation l'illumination change sensiblement. La figure 2b représente l'image des entropies de cooccurrences des 12 niveaux de gris obtenus par la procédure de multi-seuillage de la première image (figure 2a). On peut remarquer que les plus fortes valeurs (niveaux de gris clairs) de cet attribut correspondent bien à des zones d'intérêt représentatives de l'image. Les résultats d'extraction et de sélection des points d'intérêt sont présentés dans la figure 3. Pour cet exemple, parmi les 55 points extraits, seuls 32 sont localisés dans des zones à forte entropie et ainsi considérés comme étant de bons candidats. La figure 4 montre la trajectoire de suivi des points sélectionnés par les deux approches (Méthode VP et Méthode ENT). On peut noter que pour un nombre non négligeable de points (correspondant généralement à des points localisés sur des zones à faible entropie), le suivi s'arrête avant la fin de la séquence. La figure 5 montre que l'évolution du résidu global normalisé (cumul des résidus divisé par le nombre de points suivis après traitement de chaque image) en fonction du numéro de l'image dans la séquence est plus faible lorsqu'on effectue un suivi des points sélectionnés selon le critère d'entropie. Dans le même esprit, le tableau 1 récapitule les résultats obtenus. En particulier, il montre qu'un gain relatif de l'ordre de 35% sur le pourcentage des points suivis et un gain de l'ordre de 6% sur les résidus sont obtenus grâce à l'introduction du critère d'en-

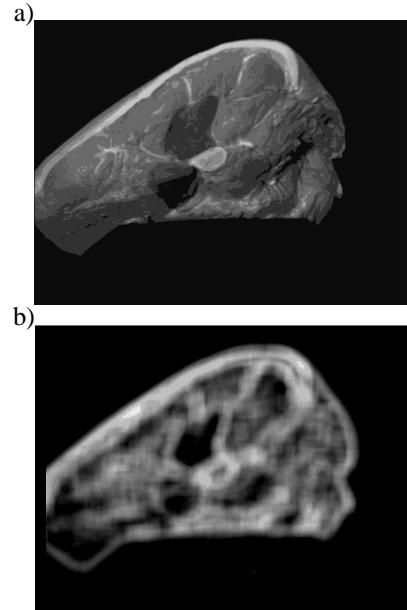


FIG. 2 – a) Résultat du multi-seuillage de la première image de la séquence PA. b) Image des entropies de cooccurrences calculées sur les niveaux de gris de l'image (a).

tropie. En d'autre terme, le classement obtenu par la méthode ENT est plus significatif de la façon dont les points sélectionnés seront suivis. Par ailleurs, en se plaçant dans un contexte d'asservissement visuel où un nombre réduit de primitives visuels de type points est nécessaire, l'expérimentation montre que 60% parmi les dix points les mieux classés selon les valeurs d'entropie sont correctement suivis, alors que seulement 20% des points le sont quand le classement est fait suivant les valeurs propres.

#### Séquence d'images synthétiques

L'approche proposée a été également essayée sur des séquences d'images synthétiques classiques dont la séquence bien connue «Yosemite» (Séquence YO, figure 6).

Les résultats issus de cette expérimentation sont donnés par les figures 7, 8, 9. Elles représentent les résultats d'extraction et de sélection des points d'intérêt, la trajectoire de suivi des points extraits suivant les deux méthodes (VP et ENT) et l'évolution du résidu global en fonction des images. Le tableau 2 résume quantitativement les résultats obtenus. Des remarques similaires à celles de l'expérimentation précédente peuvent être faites pour cette séquence. En particulier, le tableau 2 montre que, dans ce cas aussi, un résidu global meilleur est obtenu après sélection des points suivant les valeurs d'entropie et le pourcentage de points suivis est toujours plus important lorsqu'une sélection et un classement de points selon l'entropie est faite. Dans cet exemple le gain obtenu sur les résidus est de l'ordre de 9% et celui sur le pourcentage des points suivis est égale à 37%. Ici aussi, le suivi des dix premiers points met en évidence l'intérêt de la méthode ENT, 80% sont correctement suivis

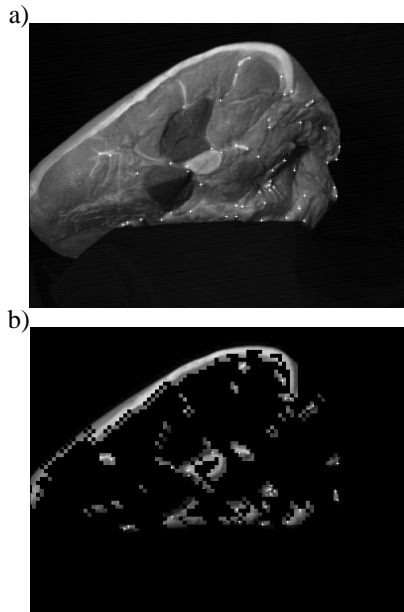


FIG. 3 – a) Résultat de l'extraction des points d'intérêt dans la première image en utilisant le critère basé sur les valeurs propres. b) Résultat de la sélection des points en utilisant le critère d'entropie (seuls les points situés dans les régions non masquées sont candidats au suivi).

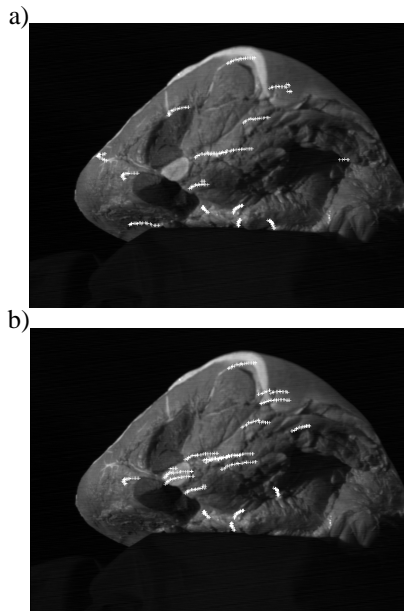


FIG. 4 – a) Trajectoires de suivi des points les mieux classés suivant les valeurs propres. b) Trajectoires de suivi des points sélectionnés suivant le critère d'entropie.

contre seulement 10% pour la méthode VP.

#### Séquence sur un objet non texturé

Une troisième série de tests concerne des séquences d'objets non texturés. Nous présentons ici des résultats obtenus sur une séquence de 100 images représentant une maquette de maison prise par une caméra en faisant un mouvement de rotation (Séquence MA, figure 10).

Les résultats issus de l'application de la méthode développée sur cette séquence sont donnés par les figures 11, 12, 13. Comme précédemment, elles représentent les résultats d'extraction et de sélection des points d'intérêt, la trajectoire de suivi des points extraits suivant les deux méthodes (VP et ENT) et l'évolution du résidu global en fonction des images. Le tableau 3 résume quantitativement les résultats obtenus. Globalement, les points sélectionnés et triés selon les valeurs d'entropie sont une fois encore mieux suivis. Les résultats obtenus montrent, ici aussi, l'intérêt de l'approche proposée dans l'élimination, dès l'image initiale, des points qui génèrent les plus forts résidus. Un gain de l'ordre de 10% sur les résidus est obtenu grâce à l'introduction de l'entropie (voir le tableau 3). Ici, 100% des dix premiers points sont correctement suivis lorsqu'ils sont classés selon les valeurs d'entropies contre 80% lorsqu'ils sont classés suivant les valeurs propres. Ce résultat permet de montrer que le gain obtenu par notre approche est faible sur une séquence où les points d'intérêt sont majoritairement des coins, primitives pour lesquelles les algorithmes d'extraction et de suivi utilisés ici sont particulièrement bien adaptés. De même, le mouvement dans cette séquence d'une image à l'autre est relativement faible et, de ce fait, le suivi n'est que rarement mis en échec.

#### Séquence d'images texturées

La dernière série de tests concerne des séquences de scènes texturées. Nous présentons ici des résultats obtenus sur la séquence connue dans la littérature sous le nom de «Flower Garden» (Séquence FG, figure 14).

Les résultats issus de l'application de la méthode développée sont donnés par les figures 15, 16, 17. Le tableau 4 résume quantitativement les résultats obtenus. Globalement, les résultats confirment, ici aussi, l'intérêt de notre approche dans l'extraction et la localisation des points susceptibles d'avoir la plus forte vraisemblance d'être correctement appariés. Un gain de l'ordre de 41% sur le pourcentage des points suivis et un gain de l'ordre de 3% sur les résidus sont obtenus grâce à l'utilisation de l'entropie. De même, parmi les dix points les mieux classés 50% sont correctement suivis lorsqu'ils le sont suivant les valeurs d'entropies, 20% seulement lorsqu'ils sont classés selon les valeurs propres.

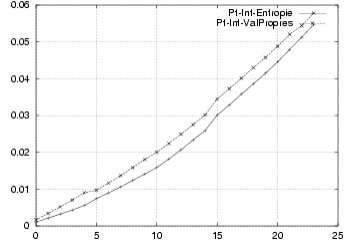


FIG. 5 – Évolution des résidus en fonction du numéro de l'image. Séquence PA.

Méth.	$Nb_{Det}$	$Nb_{Sel}$	$Nb_{Sui}$	$PPS$	$SR$
VP	55	32	11	34.38	0.057812
ENT	55	32	17	53.12	0.054744

TAB. 1 – Tableau récapitulatif pour la séquence PA.  $Nb_{Det}$ : Nombre de points extraits dans l'image initiale.  $Nb_{Sel}$ : Nombre de points sélectionnés.  $Nb_{Sui}$ : Nombre de points suivis le long de la séquence.  $PPS$ : Pourcentage des points correctement suivis.  $SR$ : Cumul des résidus normalisés.

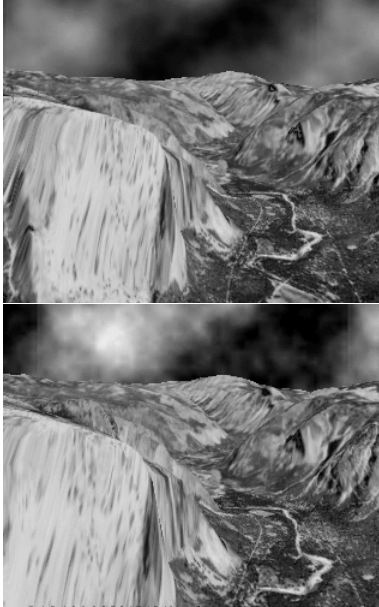


FIG. 6 – Deux images issues de la séquence Yosemite (YO).

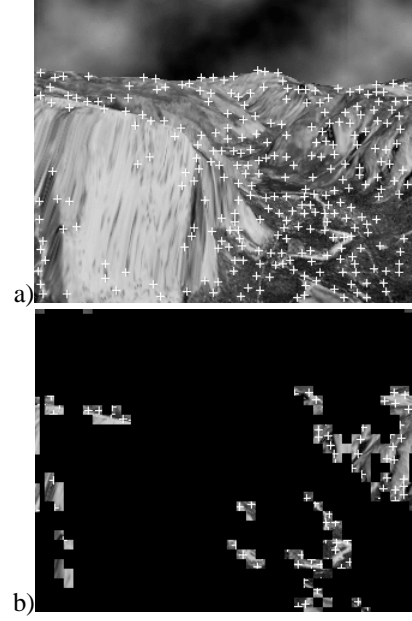


FIG. 7 – a) Résultat de l'extraction des points d'intérêt dans la première image en utilisant le critère basé sur les valeurs propres. b) Résultat de la sélection des points en utilisant le critère d'entropie.

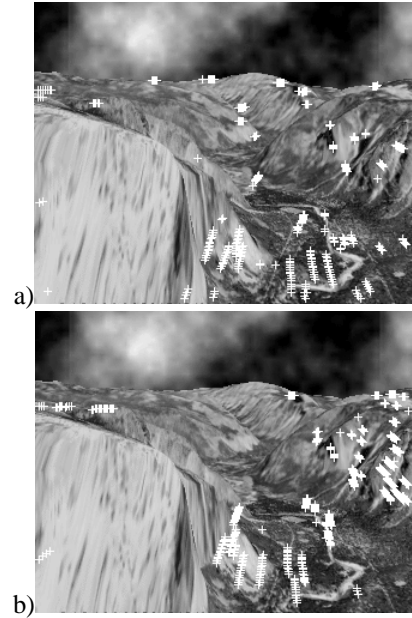


FIG. 8 – a) Trajectoires de suivi des points les mieux classés suivant les valeurs propres. b) Trajectoires de suivi des points sélectionnés suivant le critère d'entropie.

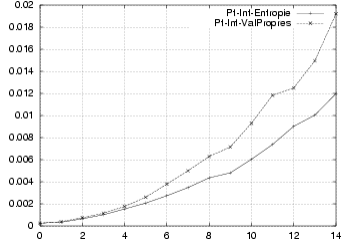


FIG. 9 – Évolution des résidus en fonction du numéro de l'image. Séquence YO.

Méth.	$Nb_{Det}$	$Nb_{Sel}$	$Nb_{Sui}$	PPS	SR
VP	281	63	11	17.46	0.019210
ENT	281	63	34	53.97	0.011980

TAB. 2 – Tableau récapitulatif pour la séquence YO.  $Nb_{Det}$ : Nombre de points extraits dans l'image initiale.  $Nb_{Sel}$ : Nombre de points sélectionnés.  $Nb_{Sui}$ : Nombre de points suivis le long de la séquence. PPS: Pourcentage des points correctement suivis. SR: Cumul des résidus normalisés.

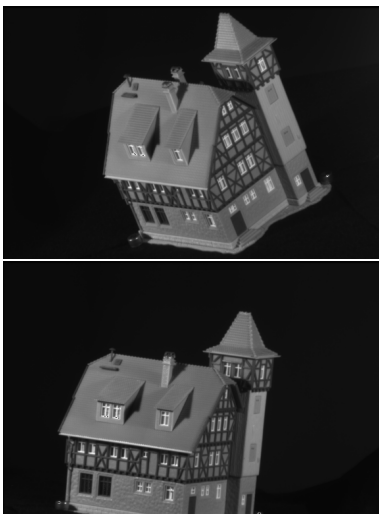


FIG. 10 – La première et la dernière image issues de la séquence MA.

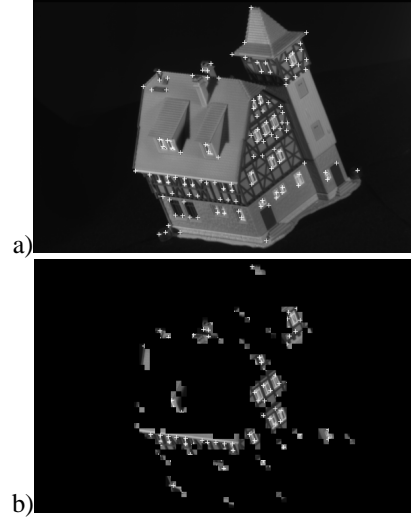


FIG. 11 – a) Résultat de l'extraction des points d'intérêt dans la première image en utilisant le critère basé sur les valeurs propres. b) Résultat de la sélection des points en utilisant le critère d'entropie.

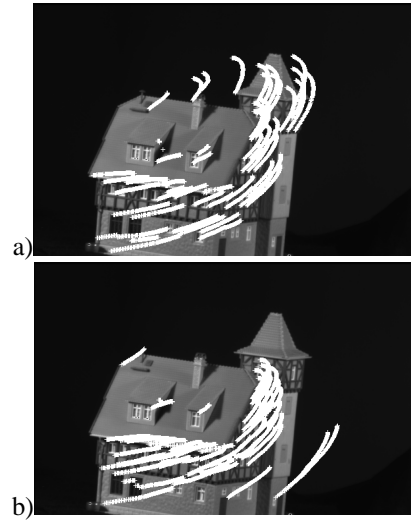


FIG. 12 – a) Trajectoires de suivi des points les mieux classés suivant les valeurs propres. b) Trajectoires de suivi des points sélectionnés suivant le critère d'entropie.

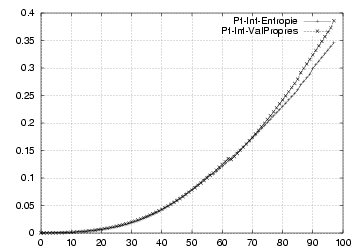


FIG. 13 – Évolution des résidus en fonction du numéro de l'image. Séquence MA.



Méth.	$Nb_{Det}$	$Nb_{Sel}$	$Nb_{Sui}$	PPS	SR
VP	123	57	47	82.46	0.386073
ENT	123	57	52	91.23	0.346008

TAB. 3 – Tableau récapitulatif pour la séquence MA.  $Nb_{Det}$ : Nombre de points extraits dans l'image initiale.  $Nb_{Sel}$ : Nombre de points sélectionnés.  $Nb_{Sui}$ : Nombre de points suivis le long de la séquence. PPS: Pourcentage des points correctement suivis. SR: Cumul des résidus normalisés.



FIG. 14 – La première et la dernière image issues de la séquence FG.

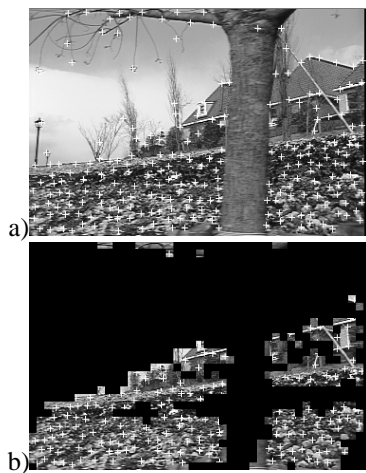


FIG. 15 – a) Résultat de l'extraction des points d'intérêt dans la première image en utilisant le critère basé sur les valeurs propres. b) Résultat de la sélection des points en utilisant le critère d'entropie.

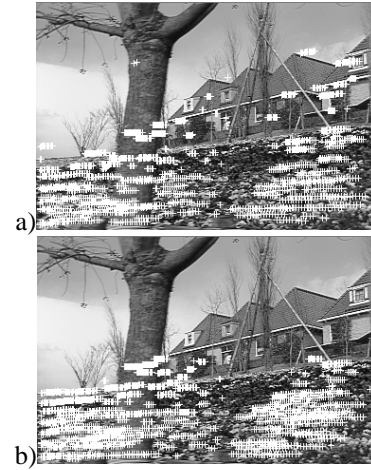


FIG. 16 – a) Trajectoires de suivi des points les mieux classés suivant les valeurs propres. b) Trajectoires de suivi des points sélectionnés suivant le critère d'entropie.

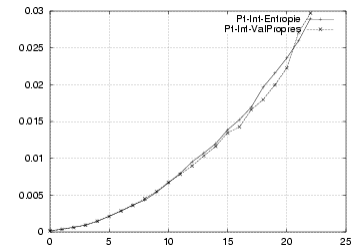


FIG. 17 – Évolution des résidus en fonction du numéro de l'image. Séquence FG.

Méth.	$Nb_{Det}$	$Nb_{Sel}$	$Nb_{Sui}$	PPS	SR
VP	231	159	23	14.47	0.029723
ENT	231	159	39	24.53	0.029006

TAB. 4 – Tableau récapitulatif pour la séquence FG.  $Nb_{Det}$ : Nombre de points extraits dans l'image initiale.  $Nb_{Sel}$ : Nombre de points sélectionnés.  $Nb_{Sui}$ : Nombre de points suivis le long de la séquence. PPS: Pourcentage des points correctement suivis. SR: Cumul des résidus normalisés.

## 7 Conclusion

Nous avons présenté une méthode de sélection de points d'intérêt dès l'image initiale d'une séquence. Elle est basée sur un critère d'entropie de cooccurrence des niveaux de gris. Cette approche permet de rejeter les points non significatifs sur la base de la richesse de la configuration locale

des pixels. Elle est en outre insensible aux changements d'intensité, de facteur d'échelle et en partie invariante aux rotations. Les résultats expérimentaux sur des images d'objets, texturés ou non, ont permis d'illustrer l'apport de notre méthode par rapport à celle de Shi-Tomasi-Kanade, classiquement utilisée. En moyenne, des gains de l'ordre de 7% sur les résidus, et de 35% sur les pourcentages des points suivis, sont obtenus grâce à l'introduction de notre critère. Une perspective de ce travail est d'améliorer la méthode proposée en la rendant moins sensible aux contours. En effet, actuellement si les points d'intérêt extraits sont situés sur des contours, ils sont conservés par le critère d'entropie sans qu'ils soient pour autant significatifs. Pour pallier à ce problème, nous comptons prendre en compte, outre le paramètre d'entropie, des informations relatives aux configurations des matrices de cooccurrence.

## Références

- [1] H. ASADA et M. BRADY. « The curvature primal sketch ». *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(1):44–54, 1986.
- [2] C.Y.C. BIE, H.C. SHEN, et D.K.Y. CHIU. « Hierarchical maximum entropy partitioning in texture image analysis ». *Pattern Recognition letters*, 14:421–429, 1993.
- [3] J. E. BYUN et T. NAGATA. « Determining the 3-D pose of a flexible object by stereo matching of curvature representations ». *Pattern Recognition*, 29(8):1297–1307, 1996.
- [4] C. COLLEWET. « Contributions à l'élargissement applicatif des asservissements visuels 2D ». Thèse de doctorat, Université de Rennes I, Rennes, Fév. 1999.
- [5] C. COLLEWET, F. CHAUMETTE, et P. LOISEL. « Image-based visual servoing on planar objects of unknown shape ». Dans *IEEE ICRA'2001*, pages 247–252, Seoul, May 21–26, 2001.
- [6] I.J. COX et S.L. HINGORANI. « An efficient implementation of Reid's Multiple Hypothesis Tracking algorithm ». *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(2):138–150, Feb. 1996.
- [7] Y. DUFOURNAUD, C. SCHMID, et R. HORAUD. « Matching Images with different resolutions ». Dans *IEEE CVPR'2000*, pages 612–618, South Carolina, USA, June 2000.
- [8] F. X. ESPIAU. « Extraction de points robustes dans des images naturelles complexes ». Dans *JJCR'2000*, Rennes, IRISA, 28–29 Sep. 2000.
- [9] J. FAN, R. WANG, L. ZHANG, et F. GAN. « Image sequence segmentation based on 2D temporal entropic thresholding ». *Pattern Recognition letters*, 17:1101–1107, 1996.
- [10] G. HAGER et P. BELHUMEUR. « Efficient Region Tracking with Parametric Models of Geometry and Illumination ». *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(10):1025–1039, Oct 1998.
- [11] R. HARALICK. « Statistical and structural approaches to textures ». *Proceedings of the IEEE*, 67(5), 1979.
- [12] C. HARRIS et M. STEPHENS. « A combined corner and edge detector ». Dans *Fourth Alvey Vision Conference*, pages 147–151, Manchester, UK, Aug. 1988.
- [13] B. K. P. HORN. « Determining lightness from an image ». *Computer Vision, graphics, and image processing*, 3(1):277–299, 1974.
- [14] B. K. P. HORN. « Image intensity understanding ». *Artificial Intelligence*, 8(2):201–231, 1977.
- [15] M. ISARD et A. BLAKE. « CONDENSATION: Conditional Propagation for Visual Tracking ». *International Journal for Computer Vision*, 29(1):5–28, 1998.
- [16] J.N. KAPUR et H.K. KESAVAN. *Entropy optimization principles with applications*. Academic Press, 1992.
- [17] C. KERMAD et K. CHEHDI. « Multi-Spectral Image segmentation : a scalar approach ». Dans *IEEE ICIP'00*, Sep. 25–28, Vancouver, British Columbia, 2000.
- [18] N. H. KIM et A. C. BOVIK. « A contour-based stereo matching algorithm using disparity continuity ». *Pattern Recognition*, 21(5):505–514, 1988.
- [19] L. KITCHEN et R. ROSENFELD. « Gray-level corner detection ». *Pattern Recognition letters*, 1:95–102, 1982.
- [20] R. D. LEVINE et M. TRIBUS. *The maximum entropy formalism*. MIT Press, 1979.
- [21] M. LHUILLIER et L. QUAN. « Robust Dense Matching Using Local and Global Geometric Constraints ». Dans *IEEE ICPR'2000*, pages 968–972, Barcelone, 2000.
- [22] B.D. LUCAS et T. KANADE. « An iterative image registration technique ». Dans *IJCAI'81*, pages 674–679, Vancouver, British Columbia, Aug. 1981.
- [23] F. MOKHTARIAN et A. MACKWORTH. « Scale-based description and recognition of planar curves and two-dimensional shapes ». *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-8(1):34–43, January 1986.
- [24] H. P. MORAVEC. « Obstacle avoidance and navigation in the real world by a seeing robot rover ». Ph. d. thesis, Stanford artificial intelligence laboratory, 1980. TR memo 340.
- [25] N.R. PAL et S.K. PAL. « Entropy: a new definition and its applications ». *IEEE Trans. on Systems, Man, and Cybernetics*, 21:1260–1270, 1991.
- [26] B.J. ROBBINS et R. OWENS. « 2D feature detection via local energy ». *Image and Vision Computing*, 15:353–368, 1997.
- [27] C. SCHMID, R. MOHRAND, et C. BAUCKHAGE. « Evaluation of Interest Point Detectors ». *International Journal of Computer Vision*, 37(2):151–172, 2000.
- [28] J. SHI et C. TOMASI. « Good features to track ». Dans *IEEE CVPR'94*, pages 593–600, Seattle, June 1994.
- [29] S. SOATTO, H. JIN, et P. FAVARO. « Real-time Feature Tracking and Outlier Rejection with Changes in Illumination ». Dans *IEEE ICCV'01*, Jul. 2001.
- [30] C. TOMASI et T. KANADE. « Detection and tracking of point features ». Technical report CMU-CS-91-132, Carnegie Mellon University, Apr. 1991.
- [31] T. TOMMASINI, A. FUSIELLO, E. TRUCCO, et V. ROBERTO. « Improving Feature Tracking with Robust Statistics ». *Pattern Analysis & Applications*, 2:312–320, 1999.
- [32] Z. ZHENG, H. WANG, et E.K. TEOH. « Analysis of gray level corner detection ». *Pattern Recognition Letters*, 20:149–162, 1999.
- [33] S.C. ZHU, Y. WU, et D. MUMFORD. « Frame: filters, random fields and maximum entropy—towards a unified theory for texture modeling ». Dans *IEEE CVPR'96*, pages 686–693, 1996.